

UNCLASSIFIED

**Defense Technical Information Center
Compilation Part Notice**

ADP014048

TITLE: Internet and Electronic Information Management

DISTRIBUTION: Approved for public release, distribution unlimited

Availability: Hard copy only.

This paper is part of the following report:

TITLE: Electronic Information Management for PfP Nations [La gestion électronique des informations pour les pays du PfP]

To order the complete compilation report, use: ADA415655

The component part is provided here to allow users access to individually authored sections of proceedings, annals, symposia, etc. However, the component should be considered within the context of the overall compilation report and not as a stand-alone technical report.

The following component part numbers comprise the compilation report:

ADP014048 thru ADP014057

UNCLASSIFIED

Click here to view PowerPoint presentation; Press Esc to exit

Internet and Electronic Information Management

Yaşar Tonta

Hacettepe University

Department of Information Management

06532 Beytepe, Ankara, Turkey

tonta@hacettepe.edu.tr

Abstract

The number of information sources accessible through the Internet are ever increasing. Billions of documents including text, pictures, sound, and video are readily available for both scholarly and every-day uses. Even libraries and information centers with sizable budgets are having difficulties in coping with this increase. More developed tools and methods are needed to find, filter, organize and summarize electronic information sources. This paper is an overview of a wide variety of electronic information management issues ranging from infrastructure to the integration of information technology and content, from personalization of information services to "disintermediation." It discusses the issues of description, organization, collection management, preservation and archiving of electronic information and outlines Davenport's "ecological model" for information management and its components, namely, strategy, politics, behavior and culture, staff, processes, and architecture.

1. Introduction

The amount of information produced in the world every year is estimated to be around two exabyte (one exabyte = one billion gigabytes, or 10^{18} bytes) (Lyman and Varian 2000). This is equal to the contents of 20 billion printed copies of the journal *The Economist* (Quantifying 2001). Information is produced in a number of different formats ranging from printed documents to the ones containing, audio, video and interactive multimedia. For instance, some 80 billion photographs, 2 billion X-rays are taken every year in the United States only. Some 610 billion e-mail messages get sent every day in the United States only. "The world's total yearly production of print, film, optical, and magnetic content would require roughly 1.5 billion gigabytes of storage. This is the equivalent of 250 megabytes per person for each man, woman, and child on earth" (Lyman and Varian 2000). As Fig. 1 shows, printed documents of all kinds constitute only .003% of the total whereas information recorded on digital formats such as personal computer disk drives and departmental servers constitute an overwhelming majority of the overall information production (Lyman and Varian 2000).

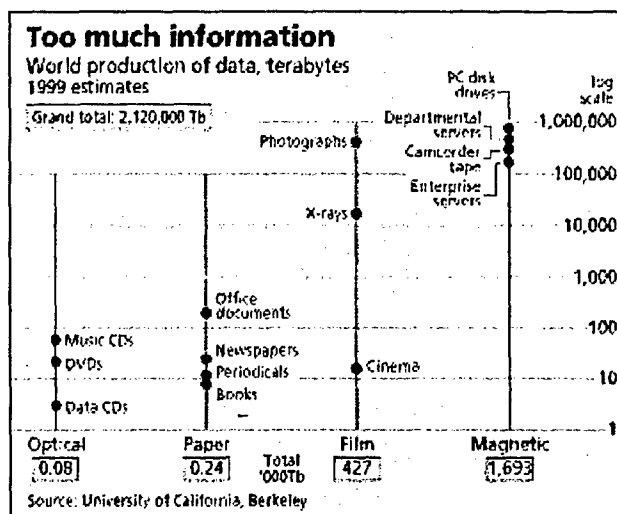


Fig. 1. World-wide information production (1999)

Source: Lyman and Varian (2000). Available:

<http://www.sims.berkeley.edu/research/projects/how-much-info/charts/charts.html>

We are faced with an enormous rate of increase of electronic information. For comparison, library collections double every 14 years whereas the *annual* growth rate for information available through the Internet is about 300%! The Library of Congress, one of the largest libraries in the world, has accumulated some 170 million items over decades. Yet the number of documents on the “surface Web” is estimated to be about 2,3 billion. These documents are open to everybody through search engines such as Google and AltaVista. Some 7,3 million new documents are added to the surface Web every day. If one includes dynamically created web pages, documents and databases available through the enterprise intranets, the number of documents goes up to 550 billion! Although accessible through the Web, such documents and databases are usually behind the firewalls and therefore not directly accessible through the regular search engines (hence called “deep Web”) (Bergman 2000).

As more information sources are born digital (or later become digital) and publicly accessible through the Internet, the relative importance of the management of information in personal, organizational and societal levels also increases tremendously. This makes the management and retrieval of information from large quantities of electronic sources all the more important. We are expected to know how to discover, find, filter, gather, organize, store, and get access to recorded information. We need to manage information successfully and be avoid “consumers” of information so as to successfully manage our professional and personal lives. Alvin Toffler (1992) warns us that the ignorants of the future are not going to be the ones who do not know how to read and write (“illiterates”) but those who do not know how to find and get access to relevant information (so called “information illiterates”). Considered by many people as the next stage of “literacy”, “information literacy” includes abilities to handle hardware and software (computers, networks, Web, etc.) used to find information as well.

Several philosophers and scholars including, among others, Plato, Bacon, and Wells have given, considerable thought to knowledge, classification, and information retrieval problems. Some discussed issues with regards to the recording, storage and retrieval of information. Plato, for instance, raises what is called “Meno’s Paradox” and draws attention to the difficulties of searching for knowledge. The following dialogue between Meno and Socrates reflects this:

“MENO. But how will you look for something when you don’t in the least know what it is? . . . even if you come right up against it, how will you know that what you have found is the thing you didn’t know?”

"SOCRATES. . . . Do you realize that . . . a man cannot try to discover either what he knows or what he does not know? He would not seek what he knows, for since he knows it there is no need of the inquiry, nor what he does not know, for in that case he does not even know what he is to look for." (*Plato's Meno* 1971: 31-32).

An interesting argument so far as it goes. Although the argument is mainly philosophical and was discussed by Plato in the context of "virtue" vs. "knowledge," searching for knowledge should not be that different from searching for information, after all. If Socrates is right, "[o]ne can never find out anything new" (*Plato's Meno* 1971: 6), nor can search for or know about anything. So, we shall not explore it further..

2. Information and communication technologies

In the past, some people were sceptical about the value of information and communication technologies (ICT) at the beginning. What follows are a few examples reflecting such scepticism:

- "Who needs this [telephone] invention? We have a lot of little boys to carry messages." Chief Engineer at the American Post Office, 1876
- "Each town may wish to have one telephone." U.S.A PTT Director General, 1886.
- "Telephone is not something that would interest millions. It is a facility for rich people; it is a commercial tool for those who could afford it." *Times*, 1902.
- "I think that as many as five computers would be sold all over the world." Thomas Watson of IBM, 1943.
- "In the future computers would weigh as little as 1.5 tons." *Popular Mechanics*, 1949.

Despite early scepticism, ICT have always played a paramount role in information processing and management. Such technologies constitute the infrastructure that "makes it possible to store, search, retrieve, copy, filter, manipulate, view, transmit, and receive information" (Shapiro and Varian 1999: 8). ICT products enable us to perform all the abovementioned activities. Shapiro and Varian (1999: 84-85) draw attention to the fact that digital technology sharply reduces *both* copying and distribution costs of information by dramatically reducing the cost of making perfect reproductions and by allowing those reproductions to be distributed quickly, easily, and cheaply. No other technology has succeeded reducing *both* reproduction and distribution costs so far.

Postman (1993: 4-5) points out that "...it is a mistake to suppose that any technological innovation has a one-sided effect. Every technology is both a burden and a blessing; not either-or, but this-and-that." Technology gets cheaper and cheaper to produce and distribute information. Yet, it increases the volume of available information and creates what is called "information overload." Reuters, the British News Agency, is producing 27.000 pages of documents per second. Users get inundated with information that they do not need, are unable to digest or simply have no time to "process." Existence of too much information creates what is called "analysis paralysis" (Waddington 1997). As the late Nobel laureate Herbert Simon put it, "a wealth of information creates a poverty of attention" (Shapiro and Varian, 1999: 6). In other words, it is not enough to provide just the content: it is much more important and difficult to "attract the eyeballs" of the potential users (Reich 2002: 42). In addition to the technology for producing and distributing information, we need technology to help simplify processing of information by the users. Otherwise, as Varian (1995: 161-162) points out, "[t]echnology for producing and distributing information is useless without some way to locate, filter, organize and summarize it."

Cost of storing and transmitting information is also sharply decreasing while the storage capacities and capabilities of various magneto-optic storage and transmission media are ever-increasing. For instance, the cost of storing data on computer hard disks went down a couple of order of magnitudes in a decade (Fig. 2). A similar trend can also be observed in the cost of transmitting data through the networks of various types (Fig. 3). Yet, such networks allow us to transmit large volumes of data in seconds. For example, a couple of years ago Nortel Networks Corporation developed a new product that "uses 160 channels, and each channel can

transmit 10 billion bits of information a second. The new system could transmit 1.6 trillion bits a second over a single optical fiber." This may not mean much to those of us who are not familiar with network speeds. To put it in context, then, Nortel points out that this transmission capacity is "enough . . . to transmit the contents of the Library of Congress across the country in 14 seconds" (Schiesel 1999).

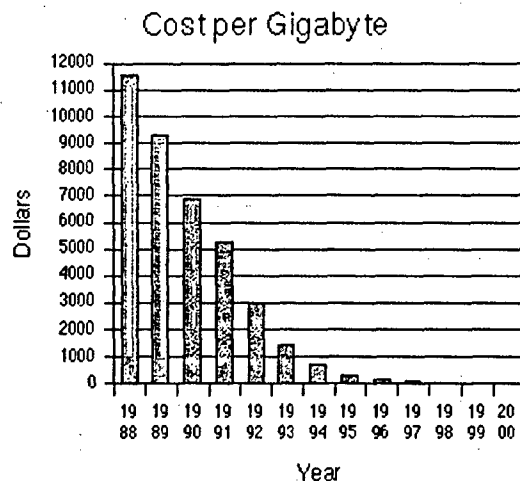


Fig. 2. Hard drive cost per gigabyte

Source: Lyman and Varian (2000). Available:

<http://www.sims.berkeley.edu/research/projects/how-much-info/charts/charts.html>

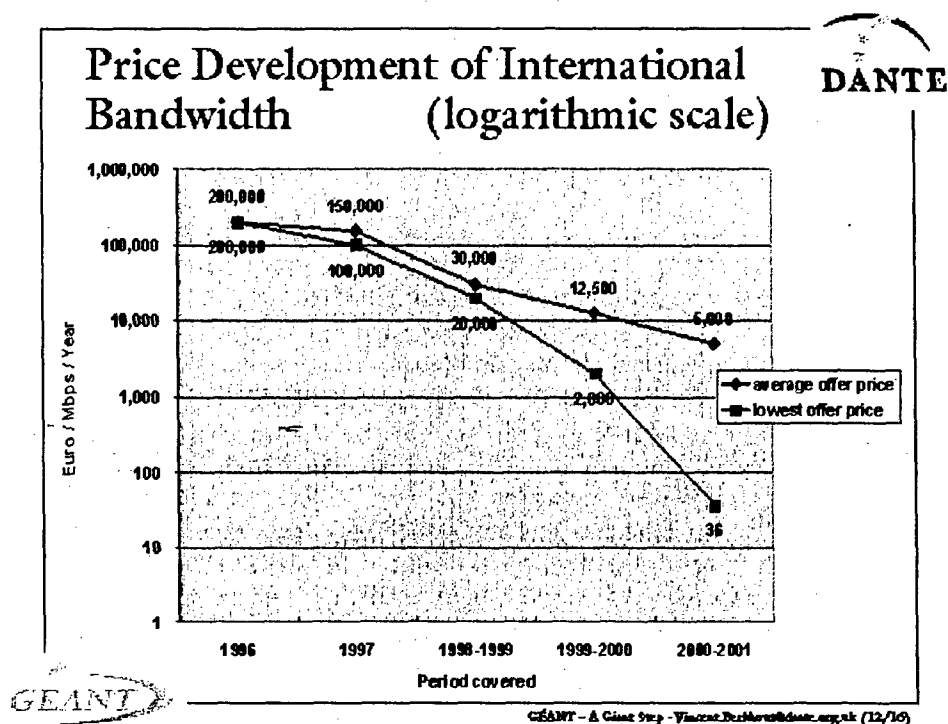


Fig. 3. Price development of international bandwidth (logarithmic scale)

Source: Berkhout (2001). Available: <http://www.dante.net/geant/presentations/vb-geant-tnc-may01/sld012.htm>.

As Varian (1995) points out, information technologies developed in recent years allow us to store all the recorded information produced by humankind throughout the centuries on a computer chip and carry it in our pockets. Moreover, such a chip containing the cumulative depository of knowledge can be implanted in our heads and used as an extension of human brain. According to George B. Dyson, the "globalization of human knowledge," as H.G. Wells prophesied some 60 years ago, is about to become a reality as many distributed databases are made available through the Internet:

The whole human memory can be, and probably in a short time will be, made accessible to every individual...This new all-human cerebrum...need not be concentrated in any one single place, it need not be vulnerable as a human head or a human heart is vulnerable. It can be reproduced exactly and fully in Peru, China, Iceland, Central Africa, or wherever else seems to afford an insurance against danger and interruption (Dyson 1997: 10-11).

Thanks to the latest developments in information technologies, the accumulated knowledge of centuries can now be distributed quickly and easily.

3. Information description and organization

Storing and transmitting large amounts of information, replicating it in different places over the globe, or even making it an extension of human brain for collective intelligence still requires speedy access to and retrieval of useful information. For this, information needs to be organized. "Information to be organized needs to be described. Descriptions need to be made of it and its physical embodiments" (Svenonius, 2000: 53). The description of information is then the *sine qua non* of both organization and, consequently, retrieval of information. Svenonius maintains that:

There is an essential difference between organizing information to compile an encyclopedic compendium of knowledge and organizing it for the purpose of information retrieval. In the former, what is ordered and arranged is the information itself; in the latter, it is the documents embodying information (such as books systematically arranged on library shelves) or their surrogates (such as catalog cards alphabetically arranged in a catalog). In the context of information retrieval, the *modus operandi* of information organization is not compilation but description (Svenonius 2000: 206).

It may at first seem an easy task to describe electronic documents (web sites, logs of discussion lists, etc.) embodying information so as to organize and later retrieve them. Yet this is one of the most difficult tasks in the Internet environment. The transient nature of Web documents sometimes makes it impossible to discover and describe information sources available through the Internet. Some Web documents are dynamically created "on-the-fly." Some ephemeral electronic documents such as meeting announcements simply disappear automatically once they fulfil their functions. In general, the average life of a Web document is estimated to be 44 days (Kahle 1997: 82-83). It is likely that some electronic documents are removed from the Web before they get noticed and described by the search engines and/or human indexers. Moreover, both the information itself (content) and its "metadata" get lost forever if it is not discovered and described before it disappears. In printed documents the information source and the metadata describing its contents are usually separated. If the printed document gets lost, one still has its metadata. This is unlike in electronic documents where both content and its metadata usually come together. Losing the document usually means losing its description as well. Although search engines are of some help in this area, they do not necessarily index all the documents available on the Web. In fact, any given search engine indexes only a fraction (e.g., 16% for Northern Light) of all the Web documents (Guernsey 1999).

We should stress the fact that describing documents is not a mechanical process. Search engines usually index documents on the basis of existence of certain keywords in the documents. They do not necessarily try to relate similar documents to one another. In fact, this is the main difference between machines and the human brain as to how they organize information. The human brain organizes information by means of what

Vannevar Bush called "associative indexing" (Bush 1945: 101-108). Even though the two pieces of information are not described with the same subject keywords, the human brain can still make connections between them.

Search engines can "describe" electronic documents simply by using some statistical techniques (e.g., the frequency of keywords). Yet, there is more to the description of documents. First, search engines use different term weighting algorithms to extract keywords from documents. Consequently, they weight and rank the same documents differently. Similarly, human indexers tend to assign different index terms to the same document. In other words, the consistency of indexing has been quite low even among professional indexers (Tonta 1991).

Second, and perhaps more importantly, it is sometimes impossible to have an agreed-upon definition of certain terms for various reasons. For instance, the term "terrorism" cannot be defined to the satisfaction of all the Nations, despite the efforts of international organizations such as the United Nations since 1970s. More currently, Nations bordering with the oil-rich Caspian Sea cannot decide whether it is a "sea" or a "lake" in view of conflicting economic stakes. If one does not have agreed-upon definitions of certain terms, how can then one describe documents satisfactorily on terrorism or the Caspian "Sea"?

Third, classification of documents, which comes after description, is another difficult issue. In his book, *Women, fire and dangerous things*, George Lakoff discovered that Australian aborigines classify women, fire and dangerous things together for some reason (Lakoff 1990). One cannot decide easily if "alternative medicine" can be categorized under the subject of health sciences, religion, philosophy or all of the above (Rosenfeld and Morville 1998: 24). We cannot even make up our minds if tomato is a vegetable or a fruit.¹

4. Collection management of electronic information

In traditional libraries, owning a book or a journal guarantees access to the contents of that information source at least by one user. If the library does not own the source, then its users have to look for it elsewhere. This is not the case for networked information sources. To provide access to such sources through the Web is as valuable as owning them on site for information centers. In light of the availability of information sources from remote locations for simultaneous use, collection development and management has become a more crucial activity in information centers. Information centers are no longer limited with their own resources to provide information services to their users. Developments of Web-based information sources and services have partly coincided with exorbitant increases of prices of printed information sources, especially scientific journals, which further encouraged information managers to try new approaches. For instance, the number of serial titles subscribed to by members of Association of Research Libraries (ARL) was decreased 7% during the period of 1986-2000, whereas the expenditures for serials has increased 192% in the same period (Fig. 4).

¹ Regarding "tomato" being a vegetable or a fruit, Rosenfeld and Morville (1998: 24), cite an interesting case (from Grady 1997) that the U.S. Supreme Court had tried in the past: "The tomato is technically a berry and thus a fruit, despite an 1893 U.S. Supreme Court decision that declared it a vegetable. (John Nix, an importer of West Indies tomatoes, had brought suit to lift a 10 percent tariff, mandated by Congress, on imported vegetables. Nix argued that the tomato is a fruit. The Court held that since a tomato was consumed as a vegetable rather than as a desert like fruit, it was a vegetable.)"

Graph 2
**Monograph and Serial Costs
 in ARL Libraries, 1986-2000**

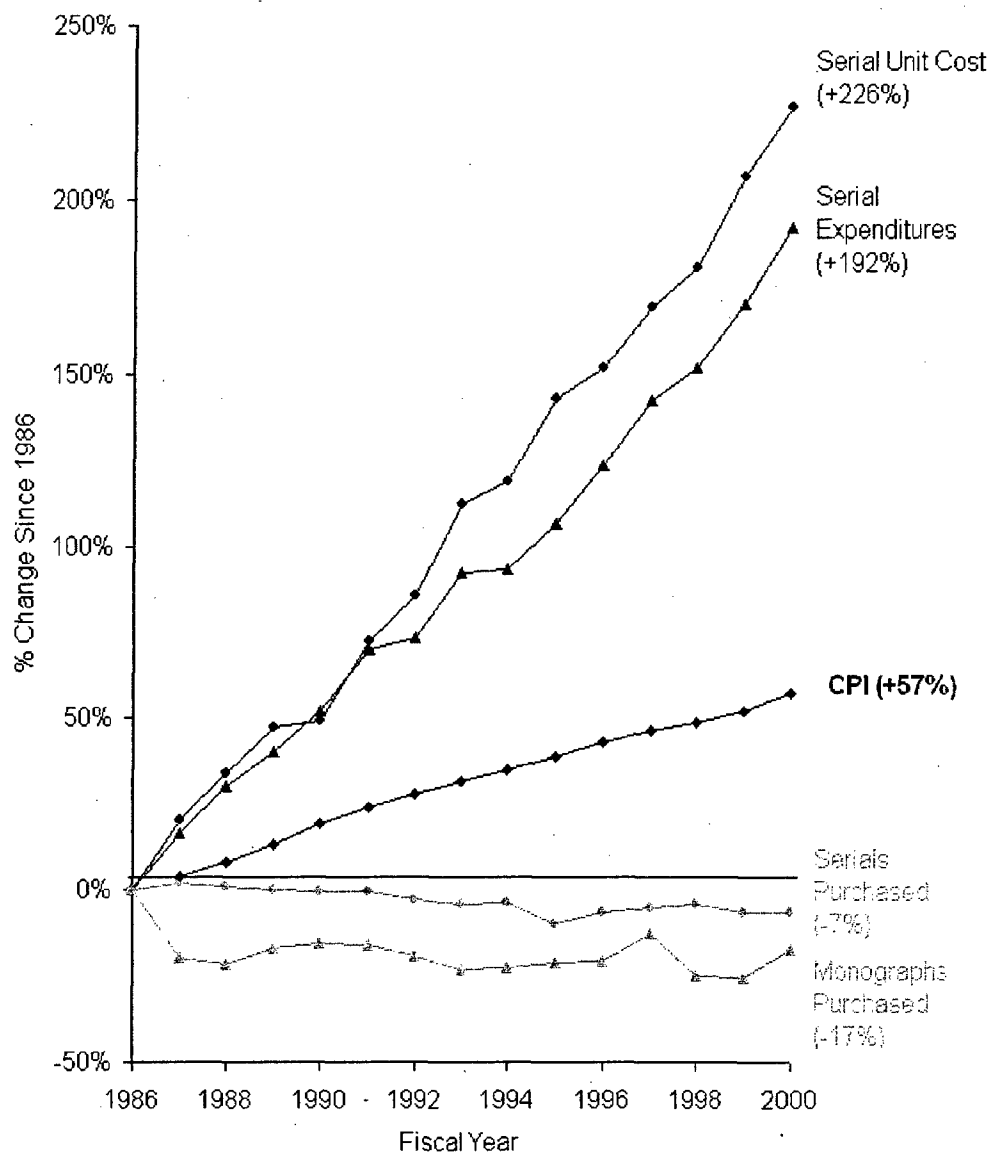


Fig. 4. Monograph and serial costs in ARL libraries, 1986-2000

Source: Kyrillidou and Young (2001, graph 2). Available:
<http://www.arl.org/stats/arlstat/graphs/2000t2.html>.

The Internet made it possible to try new financial approaches to better manage electronic information "collections." "Ownership vs. access" is one such approach.² Owning information sources "just in case" users might need them is no longer the dominant method of collection development in libraries and information centers. Instead, information centers concentrate on providing "just in time" access to electronic sources should the users need them. Ownership vs. access approach enables libraries to get access to more resources. At the same time, expenditures for processing, maintenance and storage of information sources get reduced so that more money could be spent on license fees of electronic information sources.

Vendors move from subscription-based economic models to some uncertain ones to sell information or license its use. In addition to ownership vs. access, there are a number of other approaches that could be applied in electronic collection management. Pay-per-view, transaction-based pricing, per-access charges, individual and institutional licenses or combinations thereof are among them.

It should be noted that the availability of information sources through various economic models put additional burden on collection managers. It is no longer sufficient to buy or subscribe to information sources, process, maintain and store them. Collections of information centers are not limited with what they own, maintain and archive. Separate policies of processing, maintenance, storage and usage need to be developed for different sources licensed or acquired through certain channels. Information managers need to develop policies for sources that are usually maintained and archived by other agencies. They have to develop policies for sources that they get access to through mirror sites. For instance, the electronic library of the University of California at Berkeley (<http://sunsite.berkeley.edu>) classifies electronic information sources under four groups: (1) Archived: Sources that the library own and are committed to permanent archiving as well as providing continued access; (2) Served: Sources that are maintained by the library, yet no decision has been made for their permanent archiving; (3) Mirrored: Sources that are maintained by other agencies and mirrored by the library; and (4) Linked: Sources living on remote computers, yet the library provides links to them from its own Web site (Digital, n.d.).

Naturally, responsibilities of the information centers for electronic sources under each category differ considerably from each other. For instance, in addition to institutional commitment, powerful computers, large data warehouses and bandwidth may be needed to process, store and transmit information under the first two categories (Archived and Served).

Networked information services are gradually becoming the most heavily used services in most library and information centers. What should library and information centers do to cope with this increasing demand? What types of changes, if any, should be expected in the organizational and administrative structures of library and information centers due to networked services?

It appears that the trend towards providing just in time access to more and more electronic information sources alters the use patterns of information centers. Users can conveniently consult the bibliographic databases from their own desktops located at their labs, dorms, or homes. They can download the full-texts of articles from electronic journals which their library has a license for. They can request electronic copies of books and articles from collaborating libraries or they can place online document delivery requests (if other libraries do not have access to those resources, either). These and some other automated transactions can be completed

² In fact, the trend towards access rather than ownership has been observed in other walks of life, too. The relative importance of owning property has been decreasing. For instance, people prefer to buy "experience" of different types of vacation packages rather than buying summer houses. In his book, *The age of access: how the shift from ownership to access is transforming modern life*, Jeremy Rifkin stresses that in the new economy wealth and success are measured not in terms of ownership of physical capital (plants, materials, etc.) but in terms of control of ideas in the form of intellectual and intangible capital. Ideas and talent are more important than plant and material. Rifkin points out that: "The new information-based industries—finance, entertainment, communications, business services, and education—already make up more than 25 percent of the U.S. economy. . . . The information sciences. . . are based less on ownership of physical property and more on access to valuable information, be it embedded in software or wetware" (Rifkin 2000: 52-53). Rifkin briefly touches upon the "ownership vs. access" debate taking place in the library world (p. 87-88).

without even paying a visit to the library. Just in time access increases user satisfaction and provides "instant gratification." This is reflected in the service trends in ARL libraries: the number of circulation and reference transactions has decreased for the first time in recent years while the interlibrary borrowing has almost doubled during the last decade (Kyrillidou and Young, 2001) (see Fig. 5). Consequently, library and information centers spend ever-increasing percentages of their total budgets to the provision of electronic information sources. For instance, the average percentage of the total materials budget devoted to electronic resources by U.S. research libraries was 13.5% in the academic year of 2000/2001. Some libraries (e.g., Johns Hopkins, Texas, and Arizona Universities) spent more than 20% of their total budget on electronic resources (Sewell 2001).

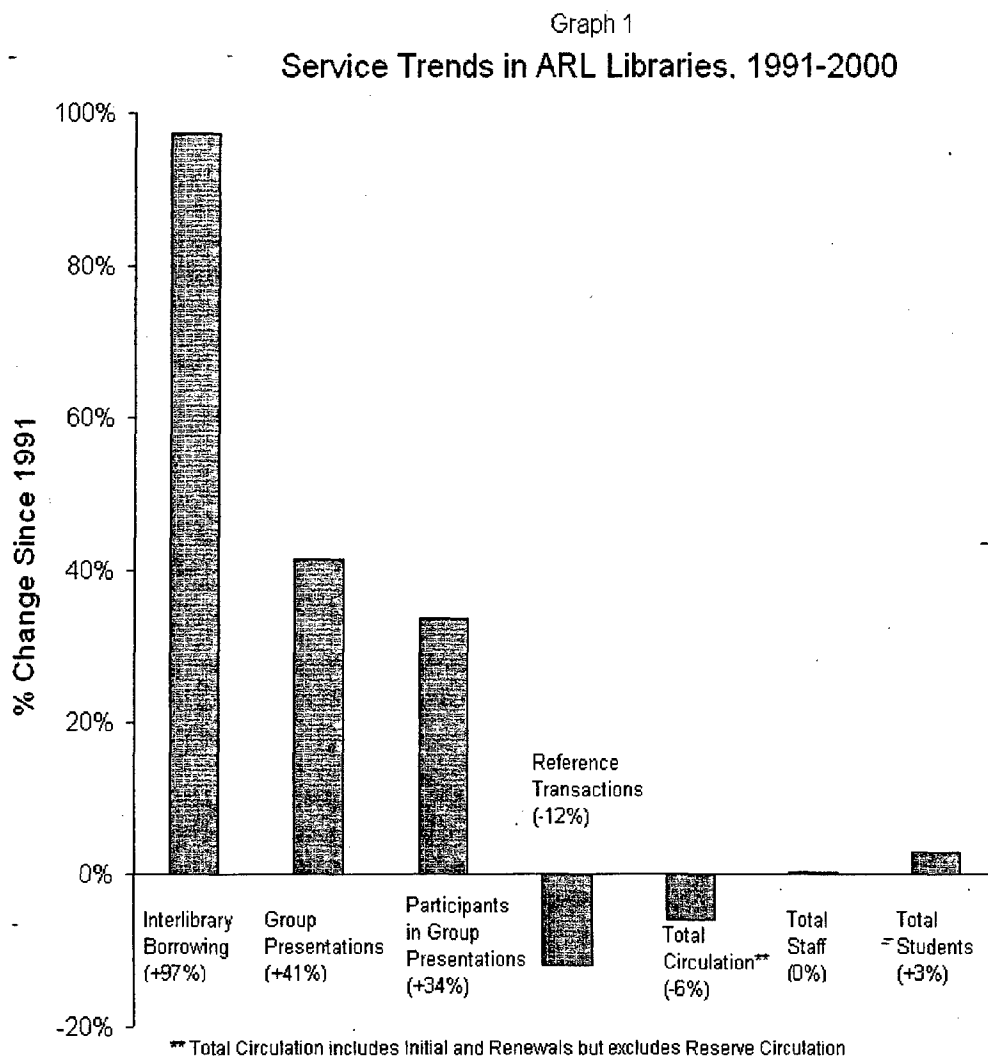


Fig. 5. Service trends in ARL libraries 1991-2000

Source: Kyrillidou and Young (2001, graph 1). Available:
<http://www.arl.org/stats/arlstat/graphs/2000t1.html>.

Information managers are faced with the challenge of providing better services with shrinking budgets. Collection managers of information centers are getting together to provide consortial access to electronic sources to get more favourable deals from information vendors. Interlibrary cooperation and coordination of resource sharing is facilitated by the network environment as it is easier for information centers to form

consortia and share electronic information sources. Although traditional resource sharing arrangements encouraged competition rather than cooperation in view of the benefits that large libraries accrued by owning research materials, this is no longer the case as small libraries and information centers can get access to information sources over the network with the same speed as the large ones can, regardless of where the sources are held. Furthermore, as indicated above, introduction of new pricing models by publishers such as licensing (rather than subscription) and access fees for electronic information sources and relatively favourable offers for consortial agreements has made the economics of cooperation more visible. This does not necessarily mean that information managers have all the wherewithal to tackle such as access management, long-term preservation and archiving of electronic information in consortial collection development schemes. Nevertheless, in terms of satisfying information needs of their users, they are in a much better position in the new environment than they were before (Tonta 2001: 292-293).

5. Mass customization and personalization of electronic information goods and services

The term "mass customization" in the title of this section may sound a bit odd as we are used to reading about "mass production" or "mass distribution" in the "mass media." As Robert R. Reich, former U.S. Secretary of Labor, points out in his book, *The future of success: working and living in the new economy*, mass production has been the pillar of the Industrial Era as it enables us to produce the same goods in large quantities in assembly lines, thereby reducing the cost of each item produced. He cites Henry Ford's famous quote "Any customer can have a car painted any color that he wants, as long as it's black" and emphasizes that "Henry Ford's assembly line lowered the cost and democratised the availability of automobile, . . . by narrowing choice" (Reich 2002: 16).

Standardized goods and services of mass production have been highly regarded in industrial societies. The futurist Alvin Toffler, on the other hand, has first mentioned the importance of "unstandardized" goods and services for the "society of the future" in 1970s in his famous book, *Future shock* (Toffler, 1970: 234-235). Developments in computer-aided manufacturing, computer and network technologies have proved Toffler right. It is now possible to produce goods exactly as you want them (which is called customization or personalization) at the best price and highest quality. Moreover, goods can be ordered from anywhere in the world as distance is no longer a constraint. In view of these three characteristics of production ("as you want them", "from anywhere", "at the best price and highest quality"), Reich labels this era as "The Age of the Terrific Deal" (Reich 2002: 13-26).

Customized production of a wide variety of goods and services in an economy is seen as an indication of a rich and complex society. As Toffler (1970: 236) emphasized: ". . . pre-automation technology yields standadization, while advanced technology permits diversity." A wide variety of customized goods and services are available in several industries: computer, automotive, textile and manufacturing industries, hotels, airlines and health services, to name just a few.

Hart (1995) defined mass customization by using two distinct definitions:

1. The visionary definition: The ability to provide customers with anything they want profitably, any time they want it, anywhere they want it, any way they want it.
2. The practical definition: The use of flexible processes and organizational structures to produce varied and often individually customized products and services at the low cost of a standardized, mass production system (cited in Mok, Stutts and Wong, 2000).

Organizational structures of companies involved in mass production and mass distribution differ from those involved in mass customization. Mass production and mass distribution is based on mechanistical organizational structures where the main objective is to produce more of the same products in large quantities and cheaper than their competitors. Administrative structure is hierarchical. Continuous development is rewarded. The training is traditional. Mass customization, on the other hand, requires dynamic organizational

structures where the main objective is to produce what the customer exactly wants. The idea is not to sell more of the same products to different customers once and make more profits but to attract more customers and keep them satisfied (as it is six times more expensive to find new customers than keeping the returning ones). Administrative structure is flattened. Customer-oriented continuous training is the norm.

The use of advanced information technologies makes it cheaper to produce personalized goods and services. Toffler points out that “. . . as technology becomes more sophisticated, the costs of introducing variations declines” (1970: 236; italics in original). This is also true of information goods and services: on-demand publishing of textbooks, readers, and newspapers, online book stores (e.g., Amazon.com), personalizing news portals (e.g., MyCNN, MyYahoo!), personalizing banking, health, education, and travel services are some of the diverse services that came into being due to advanced networking technologies.

Electronic information services are increasingly becoming personalized within the last decade. Software packages such as MyLibrary enable users to customize and personalize their electronic information environments. They define searches and identify sources that they use most frequently (including search engines, reference sources and electronic journals) and automatically get regular search results, current awareness and table of contents (TOCs) services. They get personalized document delivery and user education services. Companies try to make reference services available on every desktop computer using software (e.g., <http://www.liveperson.com>) that enables them to have reference queries answered live without even setting a foot in the library.

Electronic information services such as access to electronic journals, reference and document delivery services are gradually becoming integrated with automated library systems. For instance, to provide seamless access to electronic information sources and services, rights and privileges of each user can easily be defined depending on his/her status (i.e., student, faculty, remote user). Automated library systems can then recommend certain resources based on user's interests and privileges. Such “recommender systems” have been in use for some time by search engines and online book stores. Data gathered through the analysis of users' past interactions and transactions with the system (sites visited, books bought, etc.) are used to develop (mass) personalized information systems and recommend products and services that might conceivably be of interest to individual users. Electronic information centers also collect such valuable data about their users as well as use of their collections and services. It appears that library and information centers are reluctant to introduce recommender systems based on users and use data, presumably because they are trying to tackle issues of privacy, economics of information, and electronic payments of royalties, to name just a few.

Commercial companies producing personalized goods and services are going through a restructuring process so that they can adapt to change. They become less centralized with fewer hierarchical levels and respond to user needs more quickly. So are information centers providing electronic information services. They, too, gradually switch from centralized model of information management to the distributed one. Economic models based on centralized information management tend to produce information goods and services on the basis of “one size fits all” approach whereas models based on personalization approach users with the understanding that an ongoing relationship will be built. Personalization aims to recognize users when they use the services and provides personal help and advice if and when needed.

6. Disintermediation

Personalization of electronic information services may sound like establishing a more “personal” relationship with the users. This is not the case, however. Information about each user (his/her characteristics, habits, use patterns, etc.) is gathered by electronic means. Users sometimes provide information about themselves voluntarily by filling out forms or accepting cookies. Sometimes they have to supply information in order to get privileges of access to services. Or, sometimes aggregates of users are identified from transaction logs of system use. As a result of disappearance of face-to-face communication with users, information professionals are no longer able to “intermediate” between the users and the resources.

It appears that one of the consequences of the applications of information technology is what is called "disintermediation." Reactions towards disintermediation are mixed. Some are against it solely because they lose personal touch while others see it as an opportunity to cut costs and "reintermediate" with remote users. Take banking, for example. A face-to-face transaction carried out within the building costs banks as much as six times more of what an online transaction does. Furthermore, customers usually have to wait in the line and pay for some of those face-to-face services (e.g., money orders) while they are instantly available and free of charge if carried out through the Internet. Who will miss, then, the personal touch of, and an opportunity to exchange pleasantries with, an already overloaded teller?

Similar arguments can be put forth for personalized electronic information services as well. Face-to-face transactions cost information centers more compared to their electronic equivalents that are available through the Web. For instance, a university library comparing the costs between Web-based resources and those on its internal CD-ROM network found that average cost per search made on Web-based databases was as low as 15 cents for some databases while the average cost per search on all CD-ROMs was 15 dollars (Lindley 2000: 334). Similar comparative figures are available for the average cost of downloading an article from electronic journals as opposed to providing it through document delivery services.

Brown and Duguid (2000: 21-22) criticize the over reliance on information and point out that it leads to what they call "6-D vision." The 6-D vision consists of what authors call six "futurist-favored words" starting with "de-" or "dis-" including "disintermediation" (other five being demassification, decentralization, denationalization, despatialization, and disaggregation). They think that:

First, the evidence for disintermediation is far from clear. Organizations . . . are not necessarily becoming flatter. And, second, where it does occur, disintermediation doesn't necessarily do away with intermediaries. Often it merely puts intermediation into fewer hands with a larger grasp. The struggle to be one of those few explains several of the takeovers. . . . It also explains the "browser wars" between Netscape and Microsoft, the courtship of AT&T and Microsoft, and the continuing struggle for dominance between Internet Service Providers (ISPs). Each of these examples points not to the dwindling significance but to the continuing importance of mediation on the 'Net (as does the new term *infomediary* ...). Moreover this kind of limited disintermediation often leads to a *centralization* of control (Brown and Duguid 2000: 28, italics original).

We also witnessed such mergers and takeovers in online information industries in the last couple of years. Yet, it remains to be seen what effect, if any, they will have on the restructuring of information centers providing online information services.

7. Preservation and archiving of electronic information

Continued access to information is only possible through preservation and archiving. Preservation and archiving of electronic information sources differ significantly from that of printed sources. Preserving physical media on which information is recorded (such as books and journals) guarantees preserving the intellectual content in them. One can get access to information unless the physical medium is not damaged (Graham 1994). The first e-mail message did not survive; no "documentary" record exists for it today. Satellite observations of Brazil in the 1970s got lost as they were recorded on now obsolete tapes.

Preservation and archiving of information stored on electronic media is quite problematic. First, the life of electronic media is relatively short compared to the more traditional media such as paper and microfiche (magnetic media: 10-30 years, optical disks: 100 years, paper: 100 years, microfilm: 300 years). Second, information-bearing objects are nowadays usually "bundled" with the technology by which their content can be deciphered. In other words, one needs computer, communication and network technologies that run multimedia software to read, hear, and view electronic documents. Thus one has to buy not only the content

but also the technology as well. Furthermore, both content and the technology should be preserved so as to get access to information.

Preservation and archiving of electronic information is based on "copying." Information recorded on old media needs to be transferred to the new media media from time to time so that it will not get inaccessible due to the obsolescence of technology. New technologies do not necessarily supplant the old ones. Information published on different media (paper, microfiche, CD-ROM, etc.) will co-exist for quite some time. The copying process for preservation and archiving purposes is called "technology refreshment" or "technology migration" (Preserving 1996). Information recorded on more traditional media can also be copied onto electronic media ("digitisation"). Needless to say, migrating information from one medium to another creates several formatting problems. New versions of a software package (used with the newer technology) may sometimes not recognize the information (e.g., footnotes) that was prepared using earlier versions of the same package. Manes (1998) offers some practical guidelines for preservation and archiving of electronic information: using simple formats for copying, preserving image files without compression, using the same software for both creation and archival of electronic documents, archiving two copies of each document using good quality media, developing an archival plan before upgrading hardware and software, and, testing if new hardware is able to read the archival copies. In other words, storage, back up, refreshment and access mechanisms needed to preserve and archive electronic information should be seen as long term investments. Organizations managing electronic information need sound IT support to integrate contents with computer and communication technologies.

While the responsibility of preserving and archiving printed information rests on libraries and archives, it is not clear who is responsible for the preservation and archival of electronic information. Just as publishers were never held responsible for preservation of printed documents in the past, it is likely that they will not be relied upon for the archival copies of electronic documents, either. Some publishers refrain from even saving electronic copies of their own titles. As for-profit ventures such as publishers or "aggregators," the decision to preserve and archive is mainly shaped by commercial motives.

It is not yet clear, though, if libraries will be solely responsible for the preservation and archival of electronic information sources for various reasons. First, as we pointed out earlier, it gets cheaper to store large volumes of electronic information. Electronic information sources occupy much less space. The key question here is, of course, to decide what to preserve and what to discard. As this has been a most difficult decision to make in libraries, archives and museums, some even contemplate of preserving everything just because it is cheaper doing so. It is likely that some institutions (publishers, commercial companies, non-profit organizations such as author guilds) may wish to assume the preservation and archival of electronic information in view of economic feasibility.

Second, libraries and archives preserving printed sources do not usually get due credit for this task. They do not necessarily reap the benefits of owning those resources as the use is limited to one person only at any given time and location. The situation is quite different for institutions archiving electronic sources: they can provide access to those resources through the Internet regardless of time and location. This makes it attractive for publishers to get involved in archival business. Whereas storage and archival of printed sources consumes additional expenses, immediate access to archival copies of electronic books and journals through the publishers' Web sites generates new source of income for publishers.

Third, some publishers and non-profit organizations assume the role of long term archiving of electronic information sources published by themselves as well as other publishers. JSTOR, for instance, aims to be a reliable and long term archive of scientific journals and makes the electronic copies of those journals available through the Web. Institutions such as national libraries and universities try to become repositories of electronic copies of dissertations and technical reports. Although it remains to be seen what role information

centers are to play in the preservation and archiving of electronic information sources, the responsibility will be assumed by more than one institution in a distributed environment, as Hedstrom (1998) indicated:

...societies only allocate a small and finite amount of resources to preserving scholarly and cultural resources. And in the digital environment it seems likely that more preservation responsibilities will be distributed to individual creators, rights holders, distributors, small institutions, and other players in the production and dissemination process.

Preservation and archiving of electronic information that is based on "copying" also engenders heated discussions on issues of access and copyright for electronic information. The authenticity and integrity of copied information becomes more difficult to ascertain. More research is needed to uniquely identify digital objects for description as well as for electronic commerce and royalty payments by means of electronic copyright management systems (ECMSs). Such unresolved issues make digital preservation a "time bomb" for electronic information management (Hedstrom 1998).

8. An ecological model for electronic information management

Electronic information management can be defined as the management of information that is recorded on printed or electronic media using electronic hardware, software and networks. It includes the description of strategies, processes, infrastructure, information technology and access management requirements as well as making economic, legal and administrative policies with regards to the management of electronic information.

In his book, *Information ecology: mastering the information and knowledge environment*, Thomas H. Davenport takes a more holistic approach to information management. His approach is based on ecological paradigm which sees information in relation to its environment. What follows is a detailed review of his views on an ecological model of information management that he developed. Needless to say, his general approach to managing information is also applicable to managing electronic information.

Davenport (1997: 28) points out that most companies and organizations have done two things to manage information better: "They've applied technology to information problems, and attempted to use machine-engineering methods to turn data into something of use on computers." Davenport stresses the fact that a more holistic approach is needed and offers the "information ecology model." The following quotation summarizes his views better:

Information ecology includes a much richer set of tools than that employed to date by information engineers and architects. Information ecologists can mobilize not only architectural designs and IT but also information strategy, politics, behavior, support staff, and work processes to produce better information environments. When managers manage ecologically, they consider many avenues for achieving information objectives. They rely on disciplines of biology, sociology, psychology, economics, political science, and business strategy—not just engineering and architecture—to frame their approach to information use. And they look beyond a company's immediate information environment to the overall organizational environment—how many buildings, offices and physical locations are involved? What kind of technology is already in place? What is the current business situation?—as well as the external market environment.

... there are four key attributes of information ecology: (1) integration of diverse types of information; (2) recognition of evolutionary change; (3) emphasis on observation and description; and (4) focus on people and information behavior (Davenport 1997: 28-29).

Davenport gives the example of Amazon rain forest as a physical ecology and points out that in the overall ecology of a rain forest, there are three environments that are interconnected: treetops, shadowy world under the leaves, and soil underground. Change in one environment will result in changes in other environments.

Davenport sees three environments in any information ecology as well: information environment, organizational environment, and external environment. He thinks that:

To date, there are no practically oriented approaches that encompass all components of an information ecology—that is, how an aggregate of individuals, in a particular organization, in a particular industry affected by broader market trends, works with, thinks about, focuses on, and generally manages information. . . But description is a fundamental attribute of information ecology. And to manage ecologically, we must first understand the overall landscape in which information is used (Davenport 1997: 34).

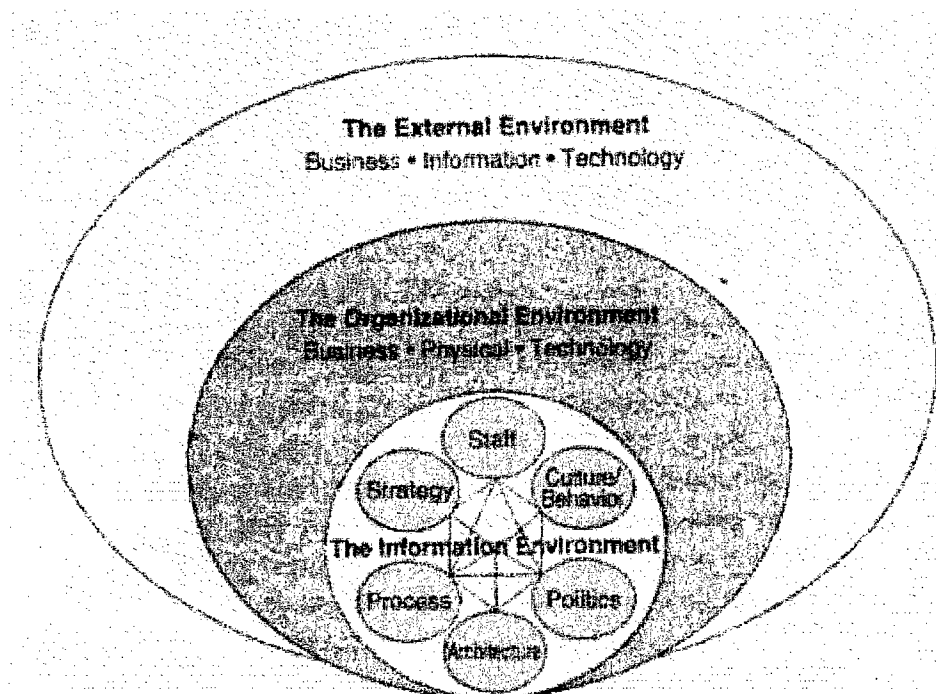


Fig. 6. An ecological model for information management
 Source: Davenport (1997: 34).

Figure 6 depicts an ecological model for information management developed by Davenport. It shows the many interconnected components of the ecological approach including the information, organizational, and external environments. What follows is a brief summary of each environment based on Davenport's (1997) account. This section owes much to his book, which I quoted and paraphrased (using mainly his words) rather liberally.

8.1. The information environment

As indicated in Fig. 6, the information environment constitutes the core of an ecological management approach and encompasses the six most critical components of information ecology—strategy, politics, behavior/culture, staff, processes, and architecture. The information environment consists of the whole set of cross-relationships among information people, strategies and policies, processes, technology, information culture and behavior.

Information Strategy: An information strategy can potentially encompass all aspects of an information ecology. The information strategy of an organization makes its high-level “information intent” explicit in an

information-pervasive world. This strategy revolves around the question, "What do we want to do with information in this organization?" An information strategy means making choices, not carving out a master plan in stone. It should outline a set of basic goals or "principles" and be flexible.

Information strategies help organizations adapt to change and make information more meaningful for the whole organization by better allocating the information resources. Davenport (1997: 49-57) describes in detail information strategies that focus on: (1) information content (gathering, analysing and acting on the most important information, be it financial, operational, or market-oriented); (2) common information (sharing common information in order to ease communications across divisions, functions and/or business processes); (3) information processes (defining processes for collection, use and disposal of information to carry out strategic goals of the organization); and (4) new information markets (using information not just internally but selling it to outside organizations, e.g., airline industries selling flight schedules to travel agents).

Information Politics: This critical component involves the power information provides and deals with the governance responsibilities for its management, control and use. Toffler (1990) observed that in most cases the way we organize information determines the way we organize people and the vice versa. Davenport (1997: 68-72) describes four model of information governance: monarchy (an "information czar" controlling most of an organization's information), federalism (a few information elements defined and managed centrally while the rest is left up to local units), feudalism (each unit manager within an organization managing his/her information environment like lords in so many separate castles), and anarchy (every individual within the organization fending for himself or herself with no centralized approach to information management initiated by the top executives). Davenport sees these four models of information control as a continuum (Fig. 7) and emphasizes the importance of matching an organization to the political structure that best suits it.

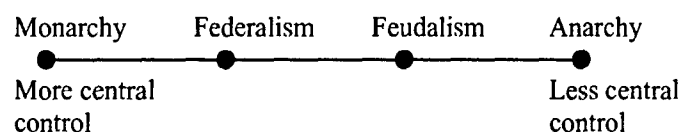


Fig. 7. The continuum of information control

Source: Davenport (1997: 69).

Information Behavior and Culture: Information behavior and culture concerns with the ways in which individuals, groups or organizations deal with information and concentrates on their approaches, attitudes and behaviors towards information. It comprises of different ways of using information (such as browsing, searching, sharing, hiding, ignoring, and making use of it). Davenport considers these two related factors most important in creating a successful information environment while, at the same time, they are the toughest to change:

All of a company's information behaviours, good or bad, make up its information culture. Particular information cultures determine how much those involved value information, share it across organizational boundaries, disclose it internally and externally, and capitalize on it in their businesses (Davenport 1997: 35).

Information Staff: Information staff consists of various professionals dealing with content (e.g., librarians, information specialists, and indexers), information technology (e.g., system designers, database administrators,

network specialists, and programmers) and other information works (e.g., management accountants, business, market, or financial analysts). Davenport rightly points out that:

People are still the best identifiers, categorizers, filterers, interpreters, and integrators of information. . . . The all-important information staff of a company handles the more valuable forms of information, such as organizational knowledge and best practices. If the information in these categories is to be of value, it must be continually pruned, restructured, interpreted, and synthesized --all tasks that computers do poorly (Davenport 1997: 35-36).

Information Processes: This component concentrates on how information work gets done. Davenport (1997: 134-152) describes a generic information management process that has four steps: determining information requirements (identifying how managers and workers make sense of their information environments); capturing information (scanning, categorizing, formatting and packaging information); distributing information (applying a combination of "pull" and "push" technologies to draw attention to the available information sources and services); and using information (assessing the information use).

Information Architecture: This component is a simple guide to the structure and location of information within organization. It is simply a set of aids that match information needs with information resources. The information architecture can be descriptive as well as prescriptive including maps, directories and standards as well as engineered models. Davenport (1997: 158-161) draws attention to the fact that most information architectures will not change culture and behaviors of information users and information staff. They may still prefer the old way of doing things.

8.2. The organizational environment

No information environment exists in and of itself. It is usually a part of the broader organizational environment and has to take into account of the organization's overall business situation, existing technology investment, and physical arrangement.

An organization's overall business situation consists of its business strategy, business processes, organizational structure and culture, and human resources. The information environment cannot fully function unless it is considered as an integral part of the overall organization and involved in the creation and development of such strategies and processes.

Existing investments of an organization on information technology determines to a great extent how the information environment carries out its responsibilities. Davenport offers the following general guidelines when investing in new technologies:

- A high degree of network interconnectedness facilitates the exchange of information in organizations;
- Knowledge and information workers require personal computers or workstations on each desktop;
- Effective information management increasingly involves providing network access to internal information repositories with many CD-based databases;
- The effective management of organizational information environments increasingly demand network management software;
- An increasing number of sophisticated software packages can help manage and distribute qualitative or document-based information in organizations;
- For external information access and communications, use of the Internet is increasingly becoming a necessity; and
- For some companies, the World Wide Web can be a new means of organizing and accessing information (Davenport 1997: 184-186).

Physical arrangement in an organization concerns with “where individuals and groups are located in relation to others with whom they work. This component also consists of the physical structures –building layouts, offices, furniture—in which people work. Finally, it includes the physical appearance and dispersal of information” (Davenport 1997: 186). Some physical arrangements facilitate information sharing while others may hinder communication of information within the organization.

8.3. The external environment

The information ecology of any organization is affected by external factors. Among these factors are government regulations, political and cultural trends in a country and in the world, business markets (customers, suppliers, competitors, regulators and public policy), technology markets (infrastructural, current-use, and innovative technologies) and information markets (buying and selling information), and the competitors’ success or failures. Such factors are beyond the control of an organization. In order for an organization to interact with the external environment, it has to:

- *adapt* to the outside world by closely monitoring government regulations and developing customer and supplier interfaces;
- *scan* that world for changes to which it must respond by identifying what external information is required, deciding where to look for it, buying and using it; and
- *mold* the outside world, through public relations and through marketing information products or services, to its own competitive advantage (Davenport 1997: 193-217).

That concludes our sketchy description of Davenport’s ecological model of information management. Although Davenport acknowledges the crucial role of technology in information management (1997: 183), he places more emphasis on other components that we outlined above. He quotes Tom Peters, the management guru, as saying “Success in information management is 5% technology, and 95% psychology” (p. 175) and thinks that:

The tools most frequently employed to design information environments derive from the fields of engineering and architecture; they rely on assumptions that may be valid when designing a building or a power generator, but rarely hold up in an organization. The sheer volume and variety of information, the multiple purposes to which it is put, and the rapid changes that take place overwhelm any rigorous attempt at central planning, design, or control (Davenport 1997: 8).

Davenport (1997: 228) believes that “No company . . . will ever achieve a true competitive advantage without adopting more human-oriented approaches to managing it. . . . It’s time to look to ourselves for the information answers.”

9. Conclusion

The Chinese proverb, “May you live in interesting times,” appears to have already been materialized: we do live in interesting times. The famous poet Paul Valery might have thought of the same thing when he said: “The future ain’t what it used to be.” This is certainly the case in the provision of electronic information services. The proliferation of electronic information products and services, increasing availability of information processing, storage and communication technologies make the jobs of information managers all the more interesting. Evolving economic paradigms based on the use, rather than ownership, of electronic information sources are a challenge for electronic information managers. Libraries and information centers are no longer “the only game in town”: they have to compete with other non-profit and for-profit providers of electronic information services. Information managers have to run their non-profit library and information centers as dynamic institutions and respond quickly to the user needs. They have to adapt to changes in the immediate information, organizational and external environments and cope with business, technology, and

market pressures. They have to cooperate with other entities within their organizations to develop more innovative information services involving the use of both internal and external sources. They have to provide the best quality personalized information services as and when their users want and deliver those services to wherever their users reside. Only then can they survive, flourish, and be "better than the best."

10. References

- Bergman, M.K. (2000 July). *The deep Web: surfacing hidden value*. (White Paper). [Online]. Available: <http://www.brightplanet.com/deepcontent/tutorials/DeepWeb/deepwebwhitepaper.pdf> [28 June 2002]. Also appeared in: *The Journal of Electronic Publishing*, 7(1), August 2001. [Online]. Available: <http://www.press.umich.edu/jep/07-01/bergman.html>. [28 June 2002].
- Berkhout, V. (2001). GÉANT: a giant step for European research. Presentation given at the TERENA Networking Conference 2001, 16 May 2001. [Online]. Available: <http://www.dante.net/geant/presentations/vb-geant-tnc-may01/> [28 June 2002].
- Brown, J.S. and Duguid, P. (2000). *The social life of information*. Boston, MA: Harvard Business School Press.
- Bush, V. (1945 July). As we may think. *Atlantic Monthly*, 176(1): 101-108. [Online]. Available: <http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm> [28 June 2002].
- Davenport, T.H. (1997). *Information ecology: mastering the information and knowledge environment*. New York: Oxford University Press.
- Digital Library SunSITE Collection and Preservation Policy. (n.d.). [Online]. Available: <http://sunsite.berkeley.edu/Admin/collection.html>. [28 June 2002].
- Dyson, G.B. (1997). *Darwin among the machines*. London: Penguin Books.
- Grady, D. (1997 July). Best bite of summer. *Self*, 19(7): 124-125.
- Graham, P.S. (1994). Intellectual preservation: electronic preservation of the third kind. *The LIBER Quarterly*, 4: 163-174.
- Guernsey, L. (1999, July 8). Seek—But on the Web, you might not find. *New York Times*, p. B8. [Online]. Available: http://www.greenwichacademy.org/studentlife/intranet/campus/tech_support/training/July8webarticle.htm [28 June 2002].
- Hart, C.H.L. (1995). Mass customization: conceptual underpinnings, opportunities and limits. *International Journal of Service Industry Management*, 6(2): 36-45.
- Hedstrom, M. (1998). Digital preservation: a time bomb for digital libraries. [Online]. Available: <http://www.uky.edu/~kiernan/DL/hedstrom.html>. [28 June 2002].
- Kahle, B. (1997 March). Preserving the Internet. *Scientific American*, 276(3): 82-83. [Online]. Available: <http://www.sciam.com/0397issue/0397kahle.html>. [10 December 2001].
- Kyrillidou, M. and Young, M. (2001). ARL statistics trends: an introduction. [Online]. Available: <http://www.arl.org/stats/arlstat/00pub/intro.html> [28 June 2002].
- Lakoff, G. (1990). *Women, fire and dangerous things: what categories reveal about the mind*. Chicago: University of Chicago Press.
- Lindley, J.A. (2000). Strategic issues in electronic librarianship. *Bilgi Dünyası*, 1(2): 330-341.
- Lyman, P. & Varian, H. (2000). How much information? [Online]. Available: <http://www.sims.berkeley.edu/how-much-info/index.html> [28 June 2002].
- Manes, S. (1998, April 7). Time and technology threaten digital archives...but with lack and diligence treasure-troves of data can be preserved. *New York Times*. Cited in: Anne Muller, e-mail message sent to EPIC-LST@NIC.SURFNET.NL [14 August 1999].
- Mok, C., Stutts, A.T. and Wong, L. (2000). Mass customization in the hospitality industry: Concepts and applications. [Online]. Available: <http://www.hotel-online.com/Neo/Trends/ChiangMaiJun00/CustomizationHospitality.html> [28 June 2002].
- Postman, N. (1993). *Technopoly: the surrender of culture to technology*. New York: Vintage Books.
- Quantifying information: Byte counters. (2001, March 26). *The Economist*, [Online]. Available: <http://www.economist.com/science/displayStory.cfm> [26 March 2001].

- Plato's Meno*. (1971). Ed. by Malcolm Brown, tr. by W.K.C. Guthrie. Indianapolis, NY: The Bobbs-Merrill Company.
- Preserving digital information*. (1996). [Online]. Report of the Task Force on Archiving Digital Information commissioned by The Commission on Preservation and Access and The Research Libraries Group. May 1, 1996. Available: <http://www.rlg.org/ArchTF/tfadi.index.htm/> or <ftp://ftp.rlg.org/pub/archtf/final-report.pdf> [28 June 2002].
- Reich, R.R. (2002). *The future of success: working and living in the new economy*. New York: Vintage Books.
- Rifkin, J. (2000). *The age of access: how the shift from ownership to access is transforming modern life*. London: Penguin Books.
- Rosenfeld, L. and Morville, P. (1998). *Information architecture for the World Wide Web*. Sebastopol, CA: O'Reilly.
- Schiesel, S. (1999, May 4). Nortel plans new product to bolster optical networks. *The New York Times*, [Online]. Available: <http://www.nytimes.com/library/tech/99/05/biztech/articles/04nortel.html> [28 June 2002].
- Sewell, R.G. (comp.) (2001, May). Library materials budget survey 2000/2001: results of the 2000/2001 library materials budget survey of the ALCTS/CMDS/chief collection development officers of large research libraries discussion group. [Online]. Available: <http://www.arl.org/scomm/lmbs/lmbs2001.html> [28 June 2002].
- Shapiro, C. and Varian, H.R. (1999). *Information rules: a strategic guide to the network economy*. Boston, MA: Harvard Business School Press.
- Svenonius, E. (2000). *The intellectual foundations of information organization*. Cambridge, MA: MIT Press.
- Toffler, A. (1970). *Future shock*. New York: Random House.
- Toffler, A. (1990). *Powershift: knowledge, wealth, and violence at the edge of the 21st century*. New York: Bantam Books.
- Tonta, Y. (1991). A study of indexing consistency between Library of Congress and British Library catalogers. *Library Resources & Technical Services*, 35(2): 177-185.
- Tonta, Y. (2001). Collection development of electronic information resources in Turkish university libraries. *Library Collections, Acquisitions & Technical Services*, 25: 291-298.
- Varian, H. (1995 September). The information economy. *Scientific American*, 273: 161-162.
- Waddington, P. (1997). Dying for information? A report on the effects of information overload in the UK and worldwide. [Online]. Available: <http://www.cni.org/regconfs/1997/ukoln-content/repor~13.html> [28 June 2002].